IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	) **MAIL STOP**
	) **APPEAL BRIEF - PATENTS**
Jerome R. Bellegarda et al.	)
	) Group Art Unit: 2168
Application No.: 10/644,815	)
	) Examiner: Mahesh H. Dwivedi
Filed: August 21, 2003	)
	) Appeal No.: _____
For: METHOD AND APPARATUS	)
FOR AUTOMATIC FILE	)
CLUSTERING INTO A DATA-	)
DRIVEN, USER SPECIFIC	)
TAXONOMY	)

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This appeal is from the decision of the Primary Examiner dated May 11, 2011, finally rejecting claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38 and 48-58, which are reproduced as the Claims Appendix of this brief.

☒   Charge $ 620 to Credit Card.

The Commissioner is hereby authorized to charge any appropriate fees under 37 C.F.R. §§ 1.17 and 41.20 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 02-4800.

Buchanan Ingersoll ⚭ Rooney PC
Attorneys & Government Relations Professionals

Table of Contents

i

I.   Real Party in Interest

Apple Inc. is the real party in interest, and is the assignee of Application No.
10/644,815.

II.   Related Appeals and Interferences

The Appellant's legal representative, or assignee, does not know of any other
appeal, interferences or judicial proceedings which will affect or be directly affected by
or have bearing on the Board's decision in the pending appeal.

III.   Status of Claims

        Claims canceled: 8, 12, 24, 29, 34 and 39-47

        Claims withdrawn from consideration but not canceled: None

        Claims pending: 1-7, 9-11, 13-23, 25-28, 30-33, 35-38 and 48-58

        Claims allowed: None

        Claims rejected: 1-7, 9-11, 13-23, 25-28, 30-33, 35-38 and 48-58

        Claims on appeal: 1-7, 9-11, 13-23, 25-28, 30-33, 35-38 and 48-58

IV.   Status of Amendments

No Amendments were filed subsequent to the final Office Action dated May 11,
2011.

V.   Summary of Claimed Subject Matter

The present application relates to automatic file clustering that enables
documents within a file system to be displayed in a semantic view, as an alternative to

displaying the documents of the file system based on user-defined storage locations of the documents.

The various files and folders in a computer system are organized in a complex hierarchy of directories, referred to as the file system. Most users start out with a reasonably principled directory structure, but as time goes by and the complexity of their file hierarchy grows, it typically becomes more difficult for them to navigate the ever-expanding portion of the file system.

Exemplary embodiments of the present invention allow navigating a file system by visualizing documents based on their content, e.g., in a semantic hierarchy. Specifically, exemplary embodiments of the present invention provide a method and apparatus for automatically clustering files in a file system and suitably displaying the resulting clusters. According to exemplary embodiments of the present invention, a semantic view option is incorporated within a graphical user interface. When invoked, this view employs a clustering and labeling algorithm that results in the creation of semantic hierarchy of all user-generated documents based on document content.

As explained above, the semantic view according to Applicants' exemplary embodiments can be incorporated into the graphical user interface as one of a number of selectable options from which the user can choose. Thus, a view might be the hierarchical tree view, as shown in Fig. 2A of the present application, in which the files are organized in accordance with their path names, i.e. the actual file system structure. To facilitate access to a particular file whose location may not be intuitive, the user can switch to the semantic view, as shown in Fig. 2B of the present application, and thereby select a file on the basis of its content, rather than its location. As a result, this semantic view option of the file system would complement a directory (folder) structured view

option based on locations of documents in the file system, and help users keep their documents in the file system in a readily usable state.

Pursuant to 37 C.F.R. §41.37(1)(c)(v), the subject matter of independent claims 1, 11, 17, 28 and 38 is cross-referenced to the specification and/or drawing figures in the following table. The following table is not to be construed as a representation that the portions of the disclosure identified below constitute the sole basis for support for the claimed subject matter.

Claim	Disclosure
1. A method of displaying files within a file system to a user in a semantic hierarchy, the method comprising the steps of:	
mapping the files in the file system into a semantic vector space;	Pages 7 and 8, paragraph 0022; Figure 3, step 301; and Figure 4
clustering the files within said space, wherein multiple threshold values that are settable to desired levels of granularity are defined, and said files are clustered based on said multiple threshold values;	Pages 7 and 8, paragraph 0022; Figure 3, step 303; and Figures 4 and 5
deriving a hierarchy of plural levels of clusters from said clustering; and	Pages 7 and 8, paragraph 0022; Figure 3, steps 305 and 307; and Figures 4 and 5
providing a user an option to selectively switch between displaying the files in a hierarchical format	Pages 13 and 14, paragraph 0036; Figures 2A and 2B

of plural levels of clusters based on said derived hierarchy, and displaying the files in a hierarchical format based on locations of the files in the file system.	
11. A non-transitory computer-readable medium containing a graphical user interface configured to display files belonging to a file system in a virtual file system with a semantic hierarchy of plural levels of clusters that is derived from semantic similarities of said files, clustering said files belonging to the file system based on multiple threshold values that are settable to desired levels of granularity, and determining a directory structure having plural levels of clusters based on the clustering determined from similarities between said files, wherein the graphical user interface provides a user an option to selectively switch between graphically displaying the determined directory structure having plural levels of clusters on a display device, and displaying the files in a hierarchical format based on locations of the files in the file system.	Pages 7 and 8, paragraph 0022; pages 13 and 14, paragraph 0036; Figures 2A and 2B show an example of a file hierarchy and a semantic hierarchy display; Figure 3 illustrates an example of creating a semantic hierarchy in a file system; Figures 4 and 5 illustrate an example of a matrix and a decomposition of documents, respectively, for clustering of the documents.

17. Non-transitory computer readable media having stored therein computer executable code for analyzing files in a file system to determine similarities in data pertaining to their content, clustering said files in the file system based on multiple threshold values that are settable to desired levels of granularity, determining a directory structure having plural levels of clusters based on the clustering determined from similarities between the files, and providing a user an option to selectively switch between displaying files in hierarchical format of plural levels of clusters based on the clustering determined from similarities between the files, and displaying the files in a hierarchical format based on locations of the files in the file system.	Pages 7 and 8, paragraph 0022; pages 13 and 14, paragraph 0036; Figures 2A and 2B show an example of a file hierarchy and a semantic hierarchy display; Figure 3 illustrates an example of creating a semantic hierarchy in a file system; Figures 4 and 5 illustrate an example of a matrix and a decomposition of documents, respectively, for clustering of the documents.
28. A computer system, comprising:	
a file system storing files;	Page 6, paragraph 0018; Figure 1, local storage disk 122;
a display device;	Page 6, paragraph 0017; Figure 1, display 104;
a processor for analyzing the content of files stored	Page 5, paragraph 0016;

in said file system to map said files into a semantic vector space, cluster the files within said space based on multiple threshold values that are settable to desired levels of granularity, and derive a hierarchy of plural levels of clusters from said clustering; and	Pages 7 and 8, paragraph 0022; Figure 1, CPU 112; and Figure 3, steps 301 and 303
a user interface which provides a user an option to selectively switch between displaying files stored in said file system in the form of said derived hierarchy of plural levels of clusters, and displaying the files in a hierarchical format based on locations of the files in the file system.	Page 6, paragraph 0017; pages 13 and 14, paragraph 0036, Figure 1, display 104, Figures 2A and 2B
38. A method of organizing a plurality of documents in a file system, comprising:	
mapping all words of the plurality of documents in the file system and the plurality of documents in a semantic vector space;	Pages 7 and 8, paragraph 0022; Figure 3, step 301; and Figure 4
generating a plurality of clusters based on the semantic similarities of the plurality of documents and multiple threshold values that are settable to desired levels of granularity;	Pages 7 and 8, paragraph 0022; Figure 3, step 303; and Figures 4 and 5
organizing the plurality of clusters into directories in a hierarchical format of plural levels of clusters; and	Pages 7 and 8, paragraph 0022; Figure 3, steps 305

	and 307; and Figures 4 and 5
providing a user an option of displaying the plurality of documents in said hierarchical format of plural levels of clusters based on a result of clustering the plurality of documents, or displaying the documents in a hierarchical format based on locations of the documents in the file system.	Pages 13 and 14, paragraph 0036; Figures 2A and 2B

VI.   Grounds of Rejection to be Reviewed on Appeal

The issue to be decided on this appeal is as follows:

1.   **Whether, under 35 U.S.C. §103(a), claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, 48, 52, 54, 56 and 58 are obvious over Bellegarda et al. (article entitled "Exploiting Latent Semantic Information in Statistical Language Modelling," hereinafter "Bellegarda") in view of Vivisimo (article entitled "Vivisimo FAQ, hereinafter "Vivisimo") and further in view of Moore et al. (U.S. Patent Application Publication No. 2004/0193621, hereinafter "Moore").**

2.   **Whether, under 35 U.S.C. §103(a), claims 49, 51, 53, 55 and 57 are obvious over Bellagarda in view of Vivisimo and Moore, and further in view of Hertz (U.S. Patent Application Publication No. 2003/0037041, hereinafter "Hertz").**

VII.   Argument

A. Claims 1-7, 9-11, 13-23, 25-28, 30-33, 35-38, 48, 52, 54, 56 and 58 are not obvious over Bellegarda in view of Vivisimo, and further in view of Moore.

1. The cited documents, even considered in combination, do not disclose all of the claim elements.

Bellegarda, Vivisimo and Moore, whether considered individually or in combination, do not enable a user to selectively switch between a hierarchical view that is <u>automatically</u> derived from a corpus of documents by mapping the files in the file system into a semantic vector space, and clustering the files, and another hierarchical view of the <u>same</u> corpus that is based upon <u>a pre-defined view</u>, such as a hierarchical view based on locations of the files. As such, Bellegarda, Vivisimo and Moore, whether considered individually or in combination, do not disclose a combination including "providing a user an option to selectively switch between displaying the files in a hierarchical format of plural levels of clusters based on said derived hierarchy, and displaying the files in a hierarchical format based on locations of the files in the file system," as recited in claim 1.

Bellegarda discloses the use of latent semantic analysis to uncover the salient semantic relationships between words and documents in a corpus. See Bellegarda: the abstract. Discrete words and documents are mapped onto a semantic vector space, in which clustering techniques can be used. *Id.* As such, Bellegarda provides a framework for automatic semantic classification of a large number of documents. *Id.* An example of the corpus is the Wall Street Journal domain. *Id.*

Vivisimo discloses organizing clustered documents in a hierarchy. See Vivisimo: page 2, under the sub-heading "What is Vivisimo doing?"

Bellegarda and Vivisimo, even considered in combination, at most disclose clustering documents in a corpus and organizing the clustered documents in a hierarchy. Bellegarda and Vivisimo, however, do not disclose enabling a user to selectively switch between a hierarchical view that is automatically derived from a corpus of documents by mapping the files in the file system into a semantic vector space, and clustering the files, and another hierarchical view of the same corpus that is based upon a pre-defined view, such as a hierarchical view based on locations of the files.

In fact, it is acknowledged, in the Office Action, that Bellegarda and Vivisimo do not disclose providing a user an option of displaying the files in a hierarchical format based on locations of the files in the file system. See the Office Action: page 4, the last full paragraph. As explained below, Moore merely discloses displaying multiple types of pre-defined views, either based on locations of the files, or manually assigned descriptions of the files stored in a database. Moore, however, does not disclose providing a display of the files in a hierarchical format based on locations of the files in the file system to a user as an option of a hierarchical view that is <u>automatically</u> derived from a corpus of documents by mapping the files in the file system into a semantic vector space, as an alternative of a pre-defined view.

Moore discloses a file organization method using virtual folders which expose regular files and folders to users in different views based on their metadata instead of the actual physical underlying file system structure on the disk. See Moore: the abstract and paragraph 0064. In Moore, the metadata include the virtual folder descriptions

stored in the virtual folder descriptions database 232, as shown in Fig. 2 of the reference.

Fig. 6 of Moore is a tree diagram of a virtual folder structure. As shown in Fig. 6, at a first level, the virtual folder 500 contains virtual folders 510, 520, and 530, corresponding to clients, contracts, and year, respectively.

Fig. 7 of Moore is a tree diagram of the virtual folder structure of Fig. 6, wherein at a second level, the virtual folder 510 further includes virtual folders 511 and 512, which correspond to contracts and year, respectively.

Fig. 8 of Moore is a tree diagram of the virtual folder structure of Fig. 7, wherein at a third level, the virtual folder 511 contains a virtual folder 513, which corresponds to a year. In other words, the contracts stack of virtual folder 511 is further filtered by year.

Moore discloses a virtual folder view as an alternative view of folders based on their locations. The virtual folder view requires virtual folder descriptions that are manually assigned indexes. See Moore, for example, paragraphs 0073-0075 and 0092 (a user re-arranges the stacks based on a property). As such, both views in Moore are pre-defined, either based on locations of the files, or the descriptions of the files stored in a database.

Moore at most can be considered as providing a pre-defined view based on locations of the files, and another pre-defined view based on the virtual folder descriptions, i.e., the virtual folder view. Moore is concerned with displaying pre-defined views on a finite amount of information in file systems. See Moore, for example, paragraphs 0002, 0073-0075 (a virtual folder descriptions database 232 includes the virtual folder descriptions). On the other hand, Bellegarda and Vivisimo are concerned with displaying a hierarchy view based on automatic classification of a potentially infinite amount of loose information, e.g., information from the Internet, or the Wall Street

Journal domain. See Vivisimo: page 1, under the sub-heading "What can benefit from Vivisimo's technology?" and Bellegarda: the abstract.

The combination of Bellegarda, Vivisimo and Moore at most discloses providing pre-defined views (e.g., a virtual folder view, or a directory view) on information if the amount of the information is finite, (e.g., file systems), and providing an automatic view if the information is an infinite amount of loose information. The documents, even considered in combination, do not disclose providing an automatic view if the amount information is finite, e.g., a file system. As such, the cited documents fail to disclose enabling a user to selectively switch between a hierarchical view that is automatically (i.e., not pre-defined) derived from a corpus of documents, and another hierarchical view of the same corpus that is based upon a pre-defined view, such as a hierarchical view based on locations of the files.

In response to Appellant's Arguments in the Amendment dated February 28, 2011, the Examiner asserts, in the Office Action dated June 29, 2011, that the Applicant's claim does not recite automatically. See the Office Action, Response to Arguments section: page 46, the last paragraph.

Appellant submits that although claim 1 does not explicitly recite "automatically" in the claim language, the recitation of the word "automatically" is not necessary since the recited steps in claim 1 clearly describe the nature of the semantic view. Claim 1 recites mapping the files in the file system into a semantic vector space; clustering the files within said space; and deriving a hierarchy of plural levels of clusters from said clustering to provide a user an option to selectively switch between displaying the files in a hierarchical format of plural levels of clusters based on said derived hierarchy, and displaying the files in a hierarchical format based on locations of the files in the file system. The hierarchical format of plural levels of clusters based on said derived

hierarchy is obtained without any need for looking up properties associated the files (e.g., locations). Therefore, such a hierarchical format of plural levels of clusters is considered as an automatically derived view, instead of a pre-defined view. Had the hierarchical format been pre-defined, claim 1 would not recite the limitation "deriving a hierarchy of plural levels of clusters from said clustering."

2. Bellegarda, Vivisimo and Moore are not properly combinable.

The Examiner asserts that it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because Moore's teaching would have allowed Bellegarda and Vivisimo to provide users the ability to toggle between virtual folder representations and physical folder representations.

Appellant submits that one with ordinary skill of the art would not apply the technique of semantic view to files in a file system. Nor is such an application suggested by either Bellegarda or Vivisimo.

The Bellegarda and Vivisimo documents focus on solutions of automatic semantic classification of documents, to avoid the cost of using manually assigned indexes. Bellegarda and Vivisimo disclose organizing a large amount of loose information, such as files from the Internet or a virtual domain. Therefore, providing an alternative view for the files that are semantically mapped, as disclosed in Bellegarda and Vivisimo, based on the locations of the files, as disclosed in Moore, would present to users a meaningless exhaustive list of the large number of files being mapped in the semantic vector space. If an Internet search is finding needles in a haystack, the above-mentioned alternative view list based on locations of a large amount of loose files is no more useful than an exhaustive list of files in the Internet.

In response to Appellant's Arguments in the Amendment dated February 28, 2011, the Examiner asserts, in the Office Action dated June 29, 2011, that applicants are arguing that the present invention gives users an option to view a meaningless exhaustive list. See the Office Action, Response to Arguments section: page 47, the first paragraph.

By making such an assertion, the Examiner has shown a lack of appreciation of the present invention. As recited in the claims, the semantic view is applied to documents in a file system. Even though the number of documents in a file system can increase, it is a finite number and a view of the documents based on their location can still be an alternative of the semantic view. In contrast, Bellegarda and Vivisimo disclose organizing a potentially infinite amount of loose information, such as files from the Internet or a virtual domain. The locations of the documents in Bellegarda and Vivisimo are widely dispersed. As such, modifying Bellegarda and Vivisimo by Moore to add a view based on locations of documents would only provide a meaningless exhaustive list of documents located in widely dispersed locations.

Appellant further submits that the virtual folder view in Moore cannot be substituted by the hierarchical view that is based on automatic semantic classification of documents in Bellegarda and Vivisimo. In Moore, items are conceptually arranged into stacks based on different properties of the items. See Moore: paragraph 0092. There are many preferences in how the items should be arranged. See Moore: paragraph 0094. The manually assigned descriptions to files and virtual folders can accommodate these different preferences by different users. For example, songs can be arranged into virtual albums. See Moore: paragraph 0092. Alternatively, songs can be arranged based on a property, e.g., a rating. To arrange the songs by albums or ratings, each of the songs must be manually assigned an album title or a rating beforehand. If the

virtual folder view in Moore is substituted by the hierarchical view that is based on automatic semantic classification of documents in Bellegarda and Vivisimo, a user cannot manually assign descriptions to the documents. As such, a user loses the ability to arrange the documents into different schema of virtual folders, such as by album titles, or by ratings. Therefore, if the virtual folder view in Moore is substituted by the hierarchical view that is based on automatic semantic classification of documents in Bellegarda and Vivisimo, the operating principle of providing users ways of arranging documents according to manually assigned descriptions to the documents in Moore would be destroyed. For these additional reasons, Bellegarda, Vivisimo and Moore are not combinable.

In view of the foregoing, the obviousness rejection of claim 1 should be reversed. The obviousness rejection of claims 2-7, 9-11, 13-23, 25-28, 30-33, 35-38, 48, 52, 54, 56 and 58 should be withdrawn for reasons similar to those for claim 1.

B. Claim 2 is not obvious over Bellegarda in view of Vivisimo, and further in view of Moore for the additional features recited therein.

With further regard to claim 2, it is asserted, in the Office Action, that Vivisimo teaches "wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system." This is clearly incorrect.

Vivisimo is concerned with organizing documents retrieved from many locations in a search result. Vivisimo discloses that "[c]lustering is done just before the user sees the search result, just in time." However, Vivisimo does not disclose that the documents are associated with a file system. Therefore, it cannot possibly disclose that the clustering is done as a background routine during the operation of a computer

associated with any particular file system. For these additional reasons, the obvious rejection of claim 2 should be reversed.

C. Claim 49, 51, 53, 55 and 57 are not obvious over Bellegarda in view of Vivisimo, and further in view of Moore and Hertz.

Claims 49, 51, 53, 55 and 57 stand rejected as being obvious over Bellagards in view of Vivisimo and Moore, and further in view of Hertz.

Hertz is not purported in the Office Action to remedy the above deficiencies of the Bellegarda, Vivisimo and Moore documents. Therefore, the obviousness rejection of the remaining pending claims should be reversed.

VIII.   Claims Appendix

See attached Claims Appendix for a copy of the claims involved in the appeal.

IX.   Evidence Appendix

None

X.   Related Proceedings Appendix

None

Respectfully submitted,

BUCHANAN INGERSOLL & ROONEY PC

Date   October 11, 2011   By: _Weiwei M. Stiltner_
                                       Weiwei Y. Stiltner
                                       Registration No. 62979

Customer No. 21839
703 836 6620

VIII. CLAIMS APPENDIX

The Appealed Claims

1.　　A method of displaying files within a file system to a user in a semantic hierarchy, the method comprising the steps of:

mapping the files in the file system into a semantic vector space;

clustering the files within said space, wherein multiple threshold values that are settable to desired levels of granularity are defined, and said files are clustered based on said multiple threshold values;

deriving a hierarchy of plural levels of clusters from said clustering; and

providing a user an option to selectively switch between displaying the files in a hierarchical format of plural levels of clusters based on said derived hierarchy, and displaying the files in a hierarchical format based on locations of the files in the file system.

2.　　The method according to claim 1, wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system.

3.　　The method according to claim 2, wherein the step of clustering the files is performed in response to the creation of a new file within the file system.

4.　　The method according to claim 1, wherein said files are text documents and said mapping is conducted on the basis of a language model.

5.   The method according to claim 4, wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector and associates each document with a vector.

6.   The method of claim 5, further including the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.

7.   The method of claim 5, wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another.

9.   The method of claim 5 further including the step of automatically labeling the clusters based on the resulting clusters.

10.   The method of claim 9 wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

11.   A non-transitory computer-readable medium containing a graphical user interface configured to display files belonging to a file system in a virtual file system with a semantic hierarchy of plural levels of clusters that is derived from semantic similarities of said files, clustering said files belonging to the file system based on multiple threshold values that are settable to desired levels of granularity, and determining a directory structure having plural levels of clusters based on the clustering determined from similarities between said files, wherein the graphical user interface provides a user an option to selectively switch between graphically

displaying the determined directory structure having plural levels of clusters on a display device, and displaying the files in a hierarchical format based on locations of the files in the file system.

13.   The non-transitory computer-readable medium according to claim 11, wherein in the graphical user interface clustering of the files is initiated by user selection.

14.   The non-transitory computer-readable medium according to claim 11, wherein in the graphical user interface clustering of the files is initiated upon creation of a new file in the file system.

15.   The non-transitory computer-readable medium according to claim 11, wherein in the graphical user interface, text files are clustered utilizing a language model and non-text files are clustered utilizing rule-based techniques.

16.   The non-transitory computer-readable medium according to claim 15, wherein in the graphical user interface, said language model comprises the LSA paradigm.

17.   Non-transitory computer readable media having stored therein computer executable code for analyzing files in a file system to determine similarities in data pertaining to their content, clustering said files in the file system based on multiple threshold values that are settable to desired levels of granularity, determining a directory structure having plural levels of clusters based on the

clustering determined from similarities between the files, and providing a user an option to selectively switch between displaying files in hierarchical format of plural levels of clusters based on the clustering determined from similarities between the files, and displaying the files in a hierarchical format based on locations of the files in the file system.

18.   The non-transitory computer-readable media of claim 17 wherein said files are text documents, and the similarities are based upon the word content of the files.

19.   The non-transitory computer-readable media of claim 18 wherein said similarities are determined in accordance with a language model, and the files are clustered in accordance with said model.

20.   The non-transitory computer-readable media of claim 19, wherein said language model comprises the LSA paradigm.

21.   The non-transitory computer-readable media of claim 19, wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector and associates each document with a vector.

22.   The non-transitory computer-readable media of claim 21, wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space.

23.   The non-transitory computer-readable media of claim 22, wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another.

25.   The non-transitory computer-readable media of claim 19, wherein said computer-executable code performs step of automatically labeling the clusters based on the resulting clusters.

26.   The non-transitory computer-readable media of claim 25, wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

27.   The non-transitory computer readable media according to claim 17, wherein the computer executable code performs the following steps:

clustering text files within the file system using semantic similarities;

clustering non-text files within the files system using rule-based techniques;

labeling the resulting clusters; and

displaying the files in a hierarchical format based on the resulting clusters and labels.

28.   A computer system, comprising:

a file system storing files;

a display device;

a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space, cluster the files within said space based on multiple threshold values that are settable to desired levels of granularity, and derive a hierarchy of plural levels of clusters from said clustering; and

a user interface which provides a user an option to selectively switch between displaying files stored in said file system in the form of said derived hierarchy of plural levels of clusters, and displaying the files in a hierarchical format based on locations of the files in the file system.

30.   The computer system of claim 28, wherein said files are text documents and said processor maps said files on the basis of a language model.

31.   The computer system of claim 30 wherein said processor constructs a matrix which associates each word in the documents with a vector and associates each document with a vector.

32.   The computer system of claim 31 wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space.

33.   The computer system of claim 31, wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another.

35. The computer system of claim 31, wherein said processor automatically labels the clusters based on the resulting clusters.

36. The computer system of claim 35 wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

37. The method according to claim 1, wherein said deriving step includes organizing the clusters into a hierarchical directory structure.

38. A method of organizing a plurality of documents in a file system, comprising:

mapping all words of the plurality of documents in the file system and the plurality of documents in a semantic vector space;

generating a plurality of clusters based on the semantic similarities of the plurality of documents and multiple threshold values that are settable to desired levels of granularity;

organizing the plurality of clusters into directories in a hierarchical format of plural levels of clusters; and

providing a user an option of displaying the plurality of documents in said hierarchical format of plural levels of clusters based on a result of clustering the plurality of documents, or displaying the documents in a hierarchical format based on locations of the documents in the file system.

48.   The method of claim 1, wherein the multiple threshold values are characteristic values of clusters from said clustering.

49.   The method of claim 48, wherein the characteristic values of the clusters are cluster variances of the clusters.

50.   The non-transitory computer-readable medium according to claim 11, wherein the multiple threshold values are characteristic values of clusters from said clustering.

51.   The non-transitory computer-readable medium according to claim 50, wherein the characteristic values of the clusters are cluster variances of the clusters.

52.   The non-transitory computer-readable media of claim 17, wherein the multiple threshold values are characteristic values of clusters from said clustering.

53.   The non-transitory computer-readable media of claim 52, wherein the characteristic values of the clusters are cluster variances of the clusters.

54.   The computer system of claim 28, wherein the multiple threshold values are characteristic values of clusters from said clustering.

55.   The computer system of claim 54, wherein the characteristic values of the clusters are cluster variances of the clusters.

56.   The method of claim 38, wherein the multiple threshold values are characteristic values of clusters from said clustering.

57.   The method of claim 56, wherein the characteristic values of the clusters are cluster variances of the clusters.

58.   The method of claim 1, further comprising providing a user an option to reorganize the files in the file system according to the derived hierarchy.

IX. EVIDENCE APPENDIX

None

X. RELATED PROCEEDINGS APPENDIX

None